Comment on esd-2021-58
Anonymous Referee #2

Referee comment on "The ExtremeX global climate model experiment: Investigating thermodynamic and dynamic processes contributing to weather and climate extremes" by Kathrin Wehrli et al., Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2021-58-RC2, 2021

Review of manuscript no. 10.5194/esd-2021-58:

"The ExtremeX global climate model experiment: Investigating thermodynamic and dynamic processes contributing to weather and climate extremes"

by

Kathrin Wehrli et al.

This paper presents the ExtremeX set of climate model experiments, where in three Earth System Models the moisture and atmospheric circulation are systematically constrained (nudged) towards observation-based values, either separately or jointly. Mean surface temperature and precipitation biases across these different experiments are evaluated, and it is found that these biases do not generally become smaller as the models are more constrained. The ExtremeX experiments are then applied to quantify the degree to which four recent strong heatwaves can be attributed to (i) sea-surface-temperature anomalies, (ii) atmospheric circulation anomalies, (iii) soil-moisture anomalies and (iv) recent climate change (from a 1979-2008 reference period to the time of the four events occurring within 2010-2015). The attribution method is then also applied to a wider set of warm spells during 2010-2015. It is found that most of the heatwaves and warm spells studied are predominantly due to circulation anomalies, with soil-moisture anomalies playing a secondary but important role, especially in subtropical and tropical regions. Contributions from sea-surface-temperature anomalies and recent climate change are typically much smaller than the other two.

The findings of this study are interesting, and it is nice to see a co-ordinated experiment across three models which lends robustness to the results, which will inform future model applications such as seasonal forecasting. I therefore recommend this study for publication in ESD subject to the comments provided below. While the presentation is generally clear, some additional investment in the introduction will make the paper more easily accessible to a wider audience. I also think that the model evaluation section would benefit from a concrete example (case study) in addition to the more general discussion provided so far. The role of the ocean in the ExtremeX setup also needs to be clarified.
General comments

1) Introduction

Having read the whole paper, and then re-read the introduction, I can follow it much better, but I think some additional explanations (and clearer signposting of contents that is already provided) would make the introduction easier to follow, especially for other readers like me who are not necessarily familiar with the predecessor papers of this study. More specifically, I recommend paying attention to the following points:

- Some key references are provided, for example in the first two paragraphs and the lead author’s own papers (line 48), but the main findings of these previous studies should be discussed in greater detail, as well as remaining knowledge gaps and which of these gaps this study aims to close.
- The focus and objectives of this study should be made clearer, especially which sort of extremes are to be studied. Line 41 rather vaguely mentions "extreme weather and climate events", whereas in the research questions it then transpires that the interest is in heatwaves/warm spells, although location, extent and duration remain unspecified. Part of my initial confusion seems to be due to the fact that there are two main purposes of this study, namely to (i) introduce the ExtremeX experiments (which I understand have a range of different possible applications) and to (ii) identify the drivers of heatwaves and warm spells, which is the specific application in this study. This distinction should be made clearer.
- Briefly motivate how to get from the conceptual distinction of dynamic and thermodynamic processes to setting up model experiments with constrained soil moisture/atmospheric circulation.

2) Validation of experiments

In section 4.2 (roughly Lines 258-283), a general discussion is provided of the issues that can arise in the constrained experiments based on tuned fully interactive models. I don’t disagree with this discussion, but it is a little unsatisfactory as it stands, and I think an example (case study), possibly in a new subsection 4.3, could help to illustrate some of these issues more clearly. A case in point already highlighted by the authors are the large summer precipitation biases seen in the MIROC5 AFSI experiment (Fig. 4) without, I believe, correspondingly large biases in clouds or evapotranspiration (Figures A3, A4). I suggest analysing this further, for example by evaluating the moisture budget (including circulation and transport) of the different experiments in a suitable study region. A possible example is WCA, where, remarkably, the precipitation bias changes sign and the RMSE increases from 0.53 to 4.3 mm/day from AISI to AFSI.

3) Role of the ocean

I am unclear about the role of the ocean in the ExtremeX setup and for the results of this study. This is illustrated in the conclusions: In Line 413, the authors say “Thus, the presented set of experiments can be used for extreme event analysis as long as the atmospheric circulation and/ or soil moisture are major drivers of the event.” This means that the role of the ocean must be small – a working assumption, or limitation of the approach. However, in Line 431 it is asserted that “The ocean was not found to have a substantial role in driving any of the events considered” – this reads like a result of this study and may be seen to be incompatible with the earlier statement. Please explain this more clearly.
Minor comments

4) Abstract

The last sentence about where soil moisture effects are important raises the expectation of a similar sentence for the circulation effects.

5) Line 19

What does “consistent” here refer to? Extreme and mean model biases? Or maybe the range of CMIP5 models?

6) Line 48

“… by validating the forcing of the atmosphere and the land for the near-surface climatology.” I did not understand this (before reading the paper). Please rephrase.

7) Line 52

Specify “overall model biases”.

8) Table 1

Provide the number of ensemble members as three comma-separated values using a specified order of models, e.g., “5,5,1” for 5 members in CESM1.2, EC-EARTH3, and 1 member in MIROC5.

9) Line 94

I suggest listing/explaining the different terminologies once upfront (forcing, constraining, nudging, relaxing) and then to stick to one choice for the remainder of the paper. “Constraining” seems to work well.

10) Line 139

Regarding the “additivity” – can this, or has this, been tested? Clarify briefly.

11) Line 141

Replace “analyses investigate” by “disentangling method determines”.

12) Line 172

Replace “The target data set” by “The prescribed target soil moisture” (if true).

13) Line 192

Explain (or omit) “non-operational”.

14) Line 206

Replace “toward observations” by “toward reanalysis” (if true). Make this distinction throughout.

15) Figure 2
This figure nicely summarises the performance for different experiments and regions!

16) Figure 4

Explain the grey areas in the caption.

17) Line 292

“… nudging the atmospheric large-scale circulation and constraining the soil moisture results ...” – The *and* seems key here as there can be substantial biases in the experiments where circulation and soil moisture are constrained individually. Please discuss if this is expected to impact the disentangling method.

18) Line 319

Previous work has suggested an important role of anomalous sea surface temperatures for the 2010 Russia heatwave (Trenberth and Fasullo 2012). This study finds that “CESM is the only model which shows a negative ocean contribution of around −7%, whereas the role of the ocean is negligible in the other models”.

Does this mean that this study contradicts Trenberth and Fasullo 2012? Is there further evidence for or against in the literature?

Such context with the existing literature should be briefly discussed – also in the conclusions and for the other three events (see also comment 1).

19) Line 376

“The spells are analysed by taking the same dates in the experiments.” – I don’t understand this.

20) Figure 7

What limits this application to events that last longer than ~2 weeks? Is this simply a question of sampling/ensemble sizes, or an inherent limitation of the disentangling method? Please discuss this briefly.

21) Figure 7

Say in the caption how the local warm season is defined.

Reference